14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

EXPERIMENTAL EVALUATION OF A LOCALIZATION ALGORITHM FOR MULTIPLE ACOUSTIC SOURCES IN REVERBERATING ENVIRONMENTS

F. Antonacci, D. Saiu, P. Russo, A. Sarti, M. Tagliasacchi, S. Tubaro

Dipartimento di Elettronica e Informazione - Politecnico di Milano - Italy

ABSTRACT

The problem of blind separation of multiple acoustic sources has been recently addressed by the TRINICON framework. By exploiting higher order statistics, it allows to successfully separate acoustic sources when propagation takes place in a reverberating environment. In this paper we apply TRINICON to the problem of source localization, emphasizing the fact that it is possible to achieve small localization errors also when source separation is not perfectly obtained. Extensive simulations have been carried out in order to highlight the trade-offs between complexity and localization error at different levels of reverberation.

1. INTRODUCTION

Consider a video-surveillance scenario where steerable cameras can be directed towards the subjects of interest. In order to enable an automatic pointing mechanism, the video sequence can be analyzed in real-time to track the relevant objects in the scene. In some applications, acoustic cues might be used together with visual ones to enhance the performance of the system by localizing the source in space.

The problem of localization of acoustic sources has been thoroughly investigated in the literature for the case of a single source and two or more receivers. Reference [1] contains a complete survey of the state-of-the-art in this area. Unfortunately, efficient techniques like GCC (Generalized Cross Correlation) [2], AED (Adaptive Eigenvalue Decomposition) [3], MCLMS (Multi-channel Least Mean Squares) [4] assume a single source setup and cannot be easily extended to multiple sources.

When two or more sources are active at the same time, the problem becomes significantly more complex as some sort of source separation must be achieved before being able to localize in space. When propagation takes place in a close environment, room reverberations cannot be neglected. In this scenario, the signals received at the microphone sensors cannot be modeled as an instantaneous mix, therefore conventional techniques based on Independent Component Analysis (ICA) fail to perform blind source separation (BSS).

Section 2 briefly summarizes the TRINICON algorithm, that has been successfully used to perform blind source separation of a non-instantaneous mix. Based on the results in [5], in this paper we investigate the application of TRINICON[5][1] as a pre-processing stage for solving the problem of localization of multiple acoustic sources. Section 3 reviews the test bed conditions used in our experiments, defining the metrics used to assess source separation and localization. Section 4 comments on the results of the simulations, showing that source localization can be achieved without perfectly separate the sources.

The work presented was developed within VISNET, a Network of Excellence (http://www.visnet-noe.org), funded by the European Commission

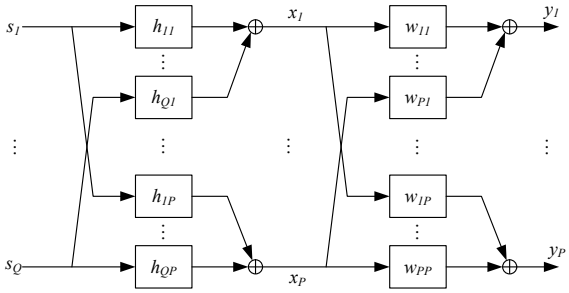

Fig. 1. Linear MIMO model for BSS

2. BACKGROUND ON TRINICON

In order to properly model room reverberations, a convolutive mixing model is better suited as it represents the signal received by each of the P microphones as the sum of delayed and filtered versions of the sources:

$$x_p(n) = \sum_{q=1}^{Q} \sum_{k=0}^{M-1} h_{qp}(k) s_q(n-k), \tag{1}$$

where Q is the number of active acoustic sources and $h_{qp}(k), k = 0, \ldots, M-1$ denote the coefficients of the finite impulse response (FIR) filter model from the q-th source to the p-th sensor. In the following, we assume that the number of source signals equals the number of sensors ($Q = P$). The goal of BSS is to find a corresponding de-mixing system according to Figure 1, where the output signals $y_q(n), q = 1, \ldots, P$ are described by:

$$y_q(n) = \sum_{p=1}^{P} \sum_{k=0}^{L-1} w_{pq}(k) x_p(n-k) \tag{2}$$

Recently, the problem of BSS for the case of multiple acoustic sources has been addressed in [6], where an iterative algorithm is used to minimize the inter-channel cross-correlation. This algorithm, originally based only on second order statistics, has been extended by the TRINICON framework [5][1]. Following the same guidelines as ICA, TRINICON efficiently exploits the non-gaussianity of the sources to improve source separation. The fundamental idea is that the sources are statistically independent and that separation is achieved when the joint inter-channel pdf of the separated signals can be factored out in the product of the pdf of each channel.

The TRINICON algorithm has been successfully used as a pre-processing stage to perform localization of multiple acoustic sources [7]. For the case of $P = Q = 2$ (two sources and two microphones), it is shown that the time difference of arrival (TDOA) expressed in

terms of number of samples can be obtained from the estimated de-mixing filters w_{pq} as follows:

$$\hat{\tau}_1 = \arg\max_n |w_{12}(n)| - \arg\max_n |w_{22}(n)| \quad (3)$$

$$\hat{\tau}_2 = \arg\max_n |w_{11}(n)| - \arg\max_n |w_{21}(n)| \quad (4)$$

The knowledge of the TDOAs allows to determine the directions of arrival of the sources with respect to the microphone array.

Equations (3) and (4) show that the information contained in the estimated de-mixing impulse responses is only partially exploited to determine the TDOAs. In other words only the position of the global maxima/minima of the impulse response are needed to achieve source localization whereas the complete impulse response is used to properly separate the sources.

Ref. [1] illustrates a specific implementation of TRINICON algorithm. The input signal is divided into non-overlapping blocks and the update of the de-mixing filters estimated at the previous step is carried out by iteratively processing each block j times. Intuitively, the separation performance tends to increase as we increase the number of iterations per block but, on the other hand, the computational cost of this implementation is proportional to j.

In this paper we elaborate on this topic, showing by means of extensive experimental simulations that partial source separation is enough when the ultimate goal is source localization. This gives rise to an interesting complexity-performance trade-off

- when little computational power is available, only source localization can be achieved
- by allowing more computational power, sources can be both separated and localized in space

It is possible to switch from one mode to the other dynamically. In fact, getting back to the video surveillance scenario, we can devise a low power tracking mode, when only source localization is carried out. In some circumstances, we might be interested in separating the sources, therefore additional power is required for a limited amount of time.

3. EXPERIMENTAL SETUP

In order to investigate TRINICON localization and separation performance we have planned a test bed evaluation. In order to simulate realistic impulse responses and acquire ground truth data, we used a beam tracing algorithm, further discussed in subsection 3.1. A typical office room, whose dimensions are $4m \times 3m \times 3m$, is assumed. Microphones are located at the center of the room and are placed $40cm$ apart. For each realization of the experiment, the positions of the two sources are randomly chosen to uniformly span the room area. Results are averaged over a set of 25 different trials.

3.1. Simulation of room reverberations

Different solutions can be adopted in order to simulate realistic impulse responses that is able to model the source-microphone channel. A first class of reverberation algorithms is based on the solution of the finite elements version of the wave equation. This class requires a significant computational effort to avoid the aliasing problem in the sampling process. The second class of simulation algorithms is based on the 'optical acoustic' solution of the wave equation. Our simulation system can be included in this second class.

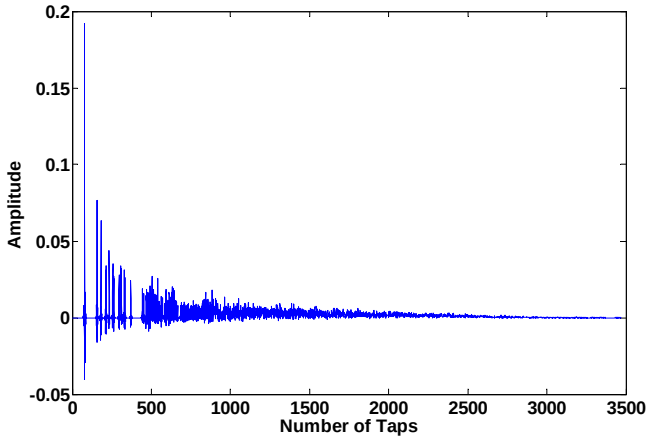

Fig. 2. Example of a simulated impulse response using $\rho = 0.8$ as reflection coefficient

In the past few years beam tracing proved to be one of the most effective algorithms in the field of room simulation. Our implementation can be considered an improvement of basic beam tracing technique (for details see [8]). First, it computes the mutual visibility between walls in a preprocessing step, without any knowledge of the source and the receiver positions. The final result of this step are diagrams (in the following 'visibility diagrams'), one for each reflecting wall, in which different regions (in the following 'visibility regions') correspond to a set of rays departing from a generic point on the considered reflector with a generic direction and impinging on a specific wall in the environment. Starting from the knowledge of walls positions in the environment we can compute visibility regions in a closed form. Once we have computed the visibility diagrams and source position is known, we can generate acoustic beams (bundles of acoustic rays) traversing the previously constructed visibility diagrams, with a traversing direction depending on the source position. Reflected beams can be built by mirroring the source over the reflector's plane (obtaining source image) and then traversing the visibility diagram associated to the mirroring reflector with a direction given by the image source position. Once we have constructed acoustic beams and receiver position is known, we can construct acoustic rays by testing the presence of receiver in previously constructed beams. With the knowledge of acoustic paths between receiver and source, we can build the desired impulse response. Several experiments showed the effectiveness of our implementation of beam tracing algorithms (see [8]).

In Figure 2, we show an impulse response (sampled at $f_s = 16kHz$) simulated in the test room, using $\rho = 0.8$ as reflection coefficient. We can distinguish the direct signal, always present in our simulation setup, followed by early reflections and late reverberations. In our simulations we used different values for the reflection coefficient in order to investigate the influence of reverberation on the accuracy of localization. Table 3.1 shows the relationship between the reflection coefficient ρ and observed reverberation time (T_{60}) in the test environment.

Once we have computed impulse responses h_{qp}, we convolve them with the original source signals, obtaining the simulated microphone signals, which are used in input to separation/localization system based on TRINICON.

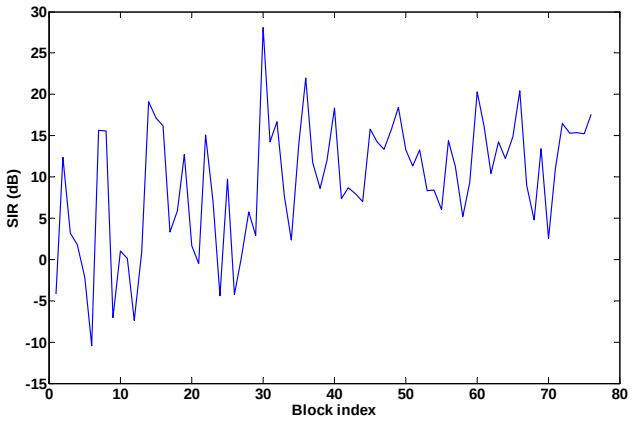

Fig. 3. Plot of SIR variation along time

3.2. Separation and localization metrics

In order to test TRINICON separation capabilities we used Signal Interference Ratio (SIR) as a separation metric. According to the notation introduced in Figure 1 SIR is computed with the following equations (respectively for the first and second signal):

$$SIR_1 = 20 * log_{10} \frac{[s_1 * (h_{11} * w_{11} + h_{12} * w_{21})]^2}{[s_2 * (h_{21} * w_{11} + h_{22} * w_{21})]^2} \quad (5)$$

$$SIR_2 = 20 * log_{10} \frac{[s_2 * (h_{22} * w_{22} + h_{21} * w_{12})]^2}{[s_1 * (h_{12} * w_{22} + h_{11} * w_{12})]^2} \quad (6)$$

Figure 3 shows an example of the SIR index variation along time. We notice the adaptive behavior of the TRINICON algorithm, as the average value of SIR tends to increase. In addition, the SIR curve variance decreases as the algorithm processes new blocks of the input signal.

Since the value of SIR depends on the input sources, it is not possible to define a fixed threshold to indicate when the algorithm converges. In alternative, we propose the following metric: first we compute the mean square root deviation σ_{conv} of the last four blocks. Then, we use a scrolling window by computing the mean square root deviation within this window σ_i, where i is the first block index belonging to the scrolling window. We conclude that the algorithm has converged at the block having index

$$i^* = \arg\min_i \sigma_i < \sigma_{conv} + 2dB \quad (7)$$

The knowledge of i^* allows to directly compute the SIR convergence time expressed in seconds.

Table 1. Relationship between reflection coefficient ρ and reverberation time (T_{60})

ρ	T_{60} (s)
0.2	0.05
0.3	0.06
0.4	0.08
0.5	0.12
0.7	0.19
0.8	0.22

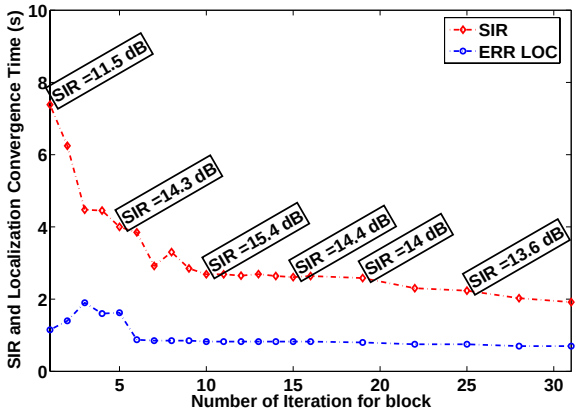

Fig. 4. SIR and localization convergence time vs. j parameter

In our experiments we assume that we know the ground truth, i.e. we know the exact location of the sources, therefore we can obtain the exact values of the TDOA expressed in a fractional number of delay samples. The algorithm successfully localize the sources when the difference between the correct TDOA and the estimated TDOA is less than half a sample. We define the TDOA convergence time as the time interval required by the algorithm to correctly localize both sources.

4. EXPERIMENTAL RESULTS

In this section we show by means of experimental results that source localization can be accomplished when sources are not perfectly separated. As mentioned in Section 2 the localization task partially exploit the estimated de-mixing impulse responses, as equations 3 and 4 are based on the positions of the maxima/minima of the de-mixing filters w_{pq}. Moreover, the separation performance of the TRINICON algorithm is influenced by the number of iterations j carried out for each block.

In the first experiment we used a reflection coefficient $\rho = 0.3$ (or equivalently, from Table 3.1 $T_{60} = 0.06s$). Figure 4 shows the convergence time expressed in seconds as a function of j. We can notice that the separation convergence time significantly decreases by increasing j. On the other hand, the effect of j on the localization convergence time is minimal. An important conclusion that can be drawn from this plot is that at a low computational cost (i.e. low values of j) we are able to localize the sources before achieving source separation. Figure 4 also shows the SIR value obtained when the algorithm achieves convergence.

Table 2. Maximum and minimum value of SIR and localization convergence time (in seconds) for different values of T_{60} (s)

	SIR		loc	
T_{60}	Max	Min	Max	Min
0.05	7.27	1.93	1.71	0.57
0.06	7.38	1.91	1.93	0.7
0.08	8.14	2.36	1.92	1.75
0.12	8.44	3.06	2.85	0.79
0.19	8.95	5.92	2.79	1.54
0.22	9.02	6.55	4.55	1.12

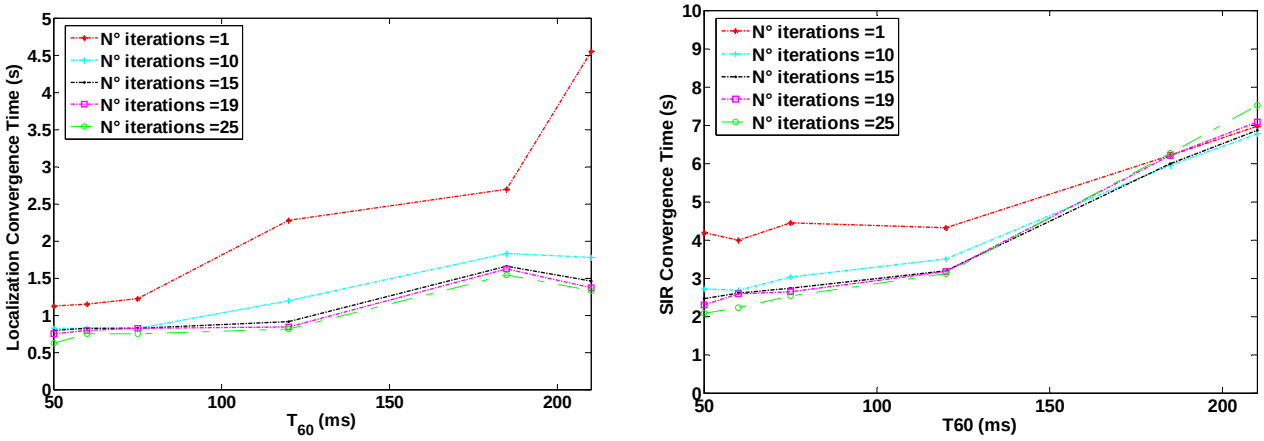

Fig. 5. Localization convergence time (left) and SIR convergence time (right) vs reverberation time T_{60} for some values of the j parameter.

We repeated the same experiment for various reflection coefficients. Table 3.2 reports the maximum and minimum values of SIR and localization convergence time. This experiment confirms that the localization task can be carried out at a lower cost with respect to the separation task.

In the second experiment we tested the sensitivity of the localization/separation algorithm to the reverberation time T_{60}. Figure 5 shows the results obtained for different values of j for the separation and localization time respectively. As expected, the convergence time tends to increase in both cases, since the estimated impulse responses are more complex. If we neglect the case of $j = 1$ the separation task 'suffers' from the increase of the reverberation time more than the localization task. As before, the latter requires to accurately estimate only the maxima/minima of the de-mixing impulse response whereas the former needs to estimate the complete response.

5. CONCLUSIONS

In this paper we have shown that localization of multiple acoustic sources can be achieved without separation as the former requires only partial knowledge of the estimated de-mixing impulse responses. Future works will extend the localization system to the problem of tracking multiple acoustic sources in reverberating environments.

6. ACKNOWLEDGMENTS

We want to acknowledge Davide Riva for helpful discussions related to this work and the support given for the simulation tests.

7. REFERENCES

[1] H. Buchner, R. Aichner, and W. Kellermann. Blind source separation for convolutive mixtures: A unified treatment. In J. Benesty and Y. Huang, editors, Audio Signal Processing for Next-Generation Multimedia Communication Systems. Kluwer Academic Publishers, Boston, Feb. 2004.

[2] J. Chen, Y. Huang, J. Banesty. "Blind source separation for convolutive mixtures: A unified treatment," In J. Benesty and Y. Huang, editors, Audio Signal Processing for Next-Generation Multimedia Communication Systems. Kluwer Academic Publishers, Boston, Feb. 2004.

[3] J. Benesty, Adaptive Eigenvalue Decomposition Algorithm for passive acoustic source localization, J. Acoust. Soc. Am., vol. 107, pp. 384−391, Jan. 2000

[4] Y. Huang and J. Benesty, Adaptive multichannel time delay estimation based on blind system identification for acoustic source localization, in Adaptive Signal Processing Applications to Real-World Problems, J. Benesty and Y. Huang, Eds., Springer, New York, 2003.

[5] H. Buchner, R. Aichner, and W. Kellermann, "TRINICON: A Versatile Framework for Multichannel Blind Signal Processing," Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), pp 889−92, vol. 3, Montreal, Canada, May 2004

[6] H. Buchner, R. Aichner, and W. Kellermann, "A Generalization of Blind Source Separation Algorithms for Convolutive Mixtures Based on Second Order Statistics," IEEE Transactions on Speech and Audio Processing, Vol. 13, No. 1, pp. 120−134, Jan. 2005.

[7] H. Buchner, R. Aichner, J. Stenglein, H. Teutsch, W. Kellermann "Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering," Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), pp 97-100, vol. 3, Philadelphia, PA, Mar 2005

[8] M. Foco, P. Polotti, A. Sarti, S. Tubaro, "Sound Spatialization Based on Fast Beam Tracing in the Dual Space", Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFx-03), London, Great Britain, Sep. 2003